

## A METHOD FOR IDENTIFICATION OF BIOLOGICALLY ACTIVE PEPTIDES AND NUCLEIC ACIDS

---

5 This invention concerns a novel method for identification of new peptides and post-translationally modified peptides as well as nucleic acids with biological activity.

### BACKGROUND OF THE INVENTION

10

During the last five years the technology for expressing, testing and identifying millions of different random peptide sequences has evolved dramatically. Such peptide libraries can be used for identification of new biologically active peptides, and therefore the technology has added an exciting and very promising new epoch to the field of drug development.

15 The known peptide library techniques can at present be divided into two fundamentally different groups: The random synthetic peptide libraries, in which the random peptides are produced chemically, and the random biosynthetic peptide libraries, in which the random peptides are encoded by partly or totally random DNA sequences and subsequently synthesized by ribosomes.

#### *The synthetic peptide libraries.*

20 Synthetic peptide libraries containing millions of peptides can be produced by combinatorial peptide chemistry and may either be synthesized in soluble form (R.A. Houghten et al. Nature, 354, 84-86, 1991) or remain immobilized on the peptide resin beads (A. Furka et al., Int. J. Peptide Protein Res. 37, 487-493, 1991; K. S. Lam et al., Nature, 354, 82-84, 1991). Using either of these approaches different receptor ligands have been isolated.

The advantage of the soluble peptide libraries compared to solid phase immobilized peptide libraries is that soluble peptides may bind more sterically unhindered to the receptors in question. In the synthetic peptide library technique proposed by Furka et al. 1991 and Lam et al. 1991, respectively, the most important improvement, on the other hand, was the approach of having only one type of peptide sequence on each bead ("One bead - one peptide"). This enables direct selection and eventually sequencing of the putative active peptide ligand on a single bead using e.g. Edman degradation. Using this technology active peptides including peptides consisting of D-amino acids or other unnatural amino acids can be identified (B. Gissel et al. J. Peptide Science. In Press, 1995).

#### *The biosynthetic peptide libraries.*

Bacteriophage expression vectors have been constructed that can display peptides on the phage surface (S.E. Cwirla et al, Proc. Natl. Acad. Sci. USA, 87, 6378-6382, 1990). Each peptide is encoded by a randomly mutated region of the phage genome, so sequencing of the relevant DNA region from the bacteriophage found to bind a receptor will reveal the amino acid sequence of the peptide ligand. A phage containing a peptide ligand is detected by repeated panning procedures which enrich the phage population for a strong receptor binding phage (J.K. Scott, TIBS, 17, 241-245, 1992).

Bacteria have also been used for expression of similar peptide libraries. The peptides can be fused to an exported protein, such as an antibody, which can be immobilized on a solid support. By screening the solid support with an appropriate soluble receptor the bacterial clone producing the putative peptide ligand can be identified (M.G. Cull et al, Proc. Natl. Acad. Sci. USA, 89, 1865-1869, 1992).

Another described method for preparing large pools of different possibly active compounds is by the use of libraries consisting of randomized ribonucleotides or deoxyribonucleotides, the so-called aptamer libraries. The  
5 aptamers are generated in *E. coli* from a plasmid vector containing randomized DNA. From these libraries structures with biological activity have been identified (L.C. Bock et al, *Nature*, 355, 564-566, 1992).

#### 10 PURPOSE OF THE INVENTION

In order to use the prior art methods for identifying biologically active peptides and nucleic acids or their  
15 respective cellular target proteins, it is necessary to possess a detailed knowledge about the molecular mechanisms involved in a certain biological process. If these mechanisms are known, it may subsequently be possible to develop antagonists or agonists of targets (receptors,  
20 enzymes, etc.) involved using said methods. The problem to be solved by the present invention is i.a. to overcome the need for said detailed knowledge.

#### SUMMARY OF THE INVENTION

25 According to the present invention the peptide sequences, or the ribonucleic acids are identified from biosynthetically expressed eukaryotic libraries containing millions of partly or totally random peptides and ribonucleic acids. In connection with this invention the term "peptide"  
30 shall be understood to comprise also a peptide sequence introduced into or fused to a larger protein (e.g. an antibody). The peptides and ribonucleic acids are synthesized by the cells from random DNA sequences which have been effectively transduced into the cells. Some of the  
35 peptides or ribonucleic acids in the library will affect important biological functions in the cells which express them. Cells which change phenotype due to the presence of such substances can be isolated, and their chemical

structure can subsequently be clarified by sequencing of the DNA which encodes them. Such peptides could possibly be used therapeutically or as lead compounds for future drug development or they can be used for identification of new target proteins which are causing the change in the biological function of the cell. Such target proteins can eventually be used in the development of drugs by e.g. conventional medicinal chemistry or synthetic peptide libraries.

Accordingly, the method of the invention comprises the following steps: (a) production of a pool of appropriate vectors each containing totally or partly random DNA sequences, (b) efficient transduction of said vectors into a number of identical, e.g. mammalian, cells in such a way that a single ribonucleic acid and possibly peptide is expressed or a limited number of different random ribonucleic acids and peptides are expressed by each cell, (c) screening of said transduced cells to see whether some of them have changed a certain phenotypic trait, (d) selection and cloning of said changed cells, (e) isolation and sequencing of the vector DNA in said phenotypically changed cells, and (f) deducing the RNA and peptide sequences from the DNA sequence.

25

#### BRIEF DESCRIPTION OF THE DRAWING

Figure 1 is a schematic drawing of a standard retroviral peptide expression vector. The plasmid form of the vector (top) carries a cytomegalovirus (CMV) promoter directing expression of a retroviral RNA (middle) with a backbone (R, U5, PBS, @, PPT, U3, and R) from Akv murine leukaemia virus and a peptide translation cassette followed by an internal ribosomal entry site (IRES) from EMC virus directing translation of a Neomycin resistance gene (Neo). The vector provirus in the target cell (below) contains a regenerated retroviral long terminal repeat - LTR (U3, R, and U5). The CMV promoter sequence provides a unique tag

for efficient initial PCR mutagenesis and amplification. The size of the vector without an inserted peptide expression cassette is 4.0 kb.

## 5 DESCRIPTION OF THE PREFERRED EMBODIMENTS

A built-in requirement of all the presently known peptide library techniques is the necessity of a detailed knowledge about the mechanisms by which a given receptor or enzyme regulates a certain phenotypic trait of the cell. This receptor or enzyme furthermore has to be available in a relatively pure form before a ligand can be selected in either of the two types of peptide libraries. When potential peptide ligands eventually have been identified, functional assays have to be performed to determine whether the ligands exert antagonistic or agonistic effects on the desired cellular phenotypic trait.

The method according to the present invention overcomes this major problem. By the present method it is thus not necessary to know the chain of mechanisms, receptors, signalling pathways, enzymes etc. which generate the phenomenon inside or on the surface of the cell since it is the resulting biological effect or phenotypic trait which is screened for. This is achieved by the following steps: (a) production of a pool of appropriate vectors each containing totally or partly random DNA sequences, (b) efficient transduction of said vectors into a number of identical eukaryotic, e.g. mammalian, cells in such a way that only a single ribonucleic acid and peptide species is expressed ("one cell - one ribonucleic acid or peptide") or optionally a limited number of different random ribonucleic acids and peptides are expressed by each cell, (c) screening of said transduced cells to see whether some of these have changed a certain phenotypic trait, (d) selection and cloning of said changed cells, (e) isolation and sequencing of the vector DNA in said phenotypically changed cells, and (f) deducing the RNA

and peptide sequences from the DNA sequence. Either the RNA or the peptide encoded by the isolated DNA sequences may be the cause of said phenotypic changes of the cell and may therefore possess biological activity.

5

The introduced peptides are expressed e.g. in the cytoplasm of biologically interesting cells, which before that were totally identical, using e.g. the retroviral vector systems described in Example 1. They are expressed from a pool of vectors containing random DNA sequences which has been constructed e.g. as described in Example 1. Using an appropriate ratio between infective retrovirus and non-infected cells only a single DNA copy derived from the pool of retrovirus vectors is introduced into each cell. The major advantage by this "one cell - one ribonucleic acid or peptide" concept is that cells which have changed phenotypically upon the introduction of peptides can be isolated by cloning and selection methods, and that the active peptide causing the phenotypic change can subsequently be identified. This is accomplished by isolating the DNA fragment encoding the peptide, e.g. by Polymerase Chain Reaction (PCR) technology, and subsequently identifying the DNA sequence.

25

During the initial screening procedure a larger number of retrovirus vectors can be introduced into each cell which enables the individual cell to express a number of different ribonucleic acids or peptides. When a phenotypically changed cell clone subsequently has been isolated all retroviral DNA in that particular clone can be isolated by PCR, and the PCR product can be used for retransfection of the packaging cells ordinarily used for virus production. The retroviral vectors isolated from these packaging cells can subsequently be used for transduction of new biologically interesting cells using the "one cell - one ribonucleic acid or peptide" concept. Finally, after a second cloning procedure the active substance can be identified as described above. Further, the

30

35

biologically active ribonucleic acid, peptide or protein isolated by this method can be utilized either as a bad compound for drug development or as an affinity ligand with the purpose of isolating and identifying the protein target responsible for the biological activity. Such target proteins are very useful tools in drug development.

The use of small vectors (below 3 kb of DNA) has the major advantage of allowing simple PCR-mutagenesis and amplification without ligation and cloning steps. Another important advantage of using small vectors is that the vectors in a pool of target cells can be amplified directly by PCR and retransfected into packaging cells, hence allowing multiple rounds of selection to remove time-consuming analysis of false positives or contaminating cells. Direct sequence analysis of the derived random plasmid clones is used to assure the randomness of the expression library. The standard vector capable of expressing the random peptide library is shown schematically in Fig. 1.

The cells which are found to have changed phenotypically upon the introduction of the random DNA sequences could alternatively have changed as a consequence of interactions with the ribonucleic acid molecule transcribed from the introduced DNA, in analogy to the described libraries consisting of randomized ribonucleotides or deoxyribonucleotides, the so-called aptamer libraries. Such ribonucleic acid molecules would therefore also possess biological activity. Furthermore, the observed effect could also be due to biological activity of carbohydrate moieties or other post-translational modifications on the expressed peptides in the cell. Using an appropriate purification-tag on the peptides in the library, it would be possible in that case to purify these and analyze the exact chemical structure of the post-translational modifications in question.

Since the efficiency of the non-viral methods commonly used for stable gene transfer into mammalian cells is very low, it would not be possible to establish a peptide expression library in mammalian cells by such methods. In addition non-viral methods generally lead to multiple integrations of DNA in the cell genome in disagreement with the "one cell one ribonucleic acid or peptide" concept. In order to achieve the necessary high efficiency single gene copy transfer a viral vector must be used. Very recently cDNA expression libraries which were constructed using retroviral vectors have been described. From such libraries cytokines and cellular growth factors have been isolated (A.J.M. Murphy et al., Proc. Natl. Acad. Sci. USA, 84, 8277-8281, 1987., B.Y. Wong et al., J. Virol., 68, 5523-31, 1994., J.R. Rayner et al., Mol. Cell. Biol., 14, 880-887, 1994). Expression of well defined peptides in transfected eukaryotic cells has also previously been established, although not using retroviral vectors (M.S. Malnati et al., Nature, 357, 702-704, 1992., E.O. Long et al., J. Immunol., 153, 1487-1494, 1994). A library of random peptides has never been expressed in mammalian cells with the purpose of identifying biologically active peptides or ribonucleic acids.

Immunology is an important biological field where the method according to the invention can be used. T cells only recognize fragments of protein antigens, and only if these are bound to MHC molecules. Two types of MHC molecules - the class I and II molecules - present antigen fragments to T cells. The peptides presented by MHC class I molecules, which are on the surface of essentially all nucleated cells, are 8-9 amino acids long and generally derived from proteins in the cytosol of the cell. These can be self-proteins, viral proteins, peptides introduced into the cytosol by transfection or tumor antigens. It is of considerable interest to be able to identify such peptides or T cell epitopes e.g. with regard to vaccine development or in immunotherapy of cancer. Identification



of such fragments is, however, a very difficult, demanding and some times impossible task requiring large amounts of affinity purified MHC molecules derived from the tumor cell in question and iterative combinations of advanced mass spectrometry, HPLC and functional T cell assays. Furthermore, most peptide antigens cannot be identified due to the presence of very low amounts of the individual peptides on the MHC molecules. In Example 2 it is demonstrated that the method according to the present invention can be used for identification of said T cell epitopes.

Cell lines expressing biologically important surface molecules can also be transduced with a random peptide library according to the invention. An example of such molecules could be the B7 co-stimulatory molecule, which is known to be important for activation of T cells, or the selectin family of proteins which are known to be involved in the homing of inflammatory cells to inflamed tissue. Cells which change phenotype (e.g. either up- or down-regulate B7) can be selected and cloned as described in Example 3. After isolation of the transduced DNA by PCR new cells can be transduced with the isolated DNA to confirm the observation. Subsequently, the peptide sequence can be deduced from the DNA sequence.

It has been described by others that specific monoclonal antibodies or F(ab) fragments can be expressed in the cytoplasm of a cell and exert a biological activity there (T.M. Werge et al., Febs Letters, 274, 193-198, 1990).

According to the present invention the wholly or partly random peptide sequence can also be introduced into the variable region of an antibody F(ab) fragment. Therefore, a library of F(ab) fragments containing random peptide sequences can be expressed in a cell clone in a way that each all express a single antibody specificity. Subsequently biologically changed cells are isolated and

cloned, and the identified intracellular F(ab) fragment can be used for purification of the target protein involved in the biological proteins. Such target protein can subsequently be used for development of drugs capable of modifying said target proteins.

The invention is illustrated by the following examples:

#### EXAMPLE 1

10

##### *Construction of an intracellular eukaryotic peptide library*

A retrovirus vector capable of expressing random peptide sequences is constructed. If the random DNA sequences used in the vector were produced using conventional random oligonucleotide synthesis a large number of stop codons inevitably would be introduced. Furthermore, due to the degeneracy of the genetic code an uneven distribution of the encoded amino acids would be the result. We avoid this by producing the random DNA sequences by random codon synthesis: An appropriate amount of resin used for solid phase oligonucleotide synthesis (optionally already containing a DNA sequence corresponding to an appropriate vector cloning site) is divided into 20 different portions. On portion no. 1 a conventional solid phase chemical synthesis of three bases corresponding to a codon encoding the amino acid, alanine, is performed. On portion no. 2 a codon encoding cysteine is synthesized and so forth. When each of the 20 portions have been coupled with codons corresponding to each of the natural amino acids, all portions are mixed and divided again into 20 equally sized resin portions. The codon synthesis is then repeated again, and the whole procedure is repeated until the desired randomized DNA sequence has been synthesized - e.g. corresponding to 6-10 random amino acids. This can also be achieved by using blocked and protected trinucleotide phosphoramidites encoding the 20

natural amino acids in a total random oligonucleotide synthesis (J. Sondek et al., Proc. Natl. Acad. Sci. USA, 89, 3581-5, 1992). Finally, other appropriate vector cloning sites can be synthesized on all the oligonucleotides before they eventually are cleaved from the resin.

The pool of random synthetic oligonucleotides can be used to generate a pool of vectors with random sequences in appropriate positions either by restriction cleavage and ligation or, preferentially, to avoid inefficient ligation steps, by PCR-mutagenesis. The procedures follow the principles of site-directed PCR-mediated mutagenesis (S. Perrin et al., Nucl. Acids Res. 18, 7433-38, 1990), but the methodology has been adapted to deliver a complex mixture of products. Briefly, the random oligonucleotides carrying vector sequences flanking the random sequence are used as primers in a PCR reaction together with a unique terminal vector primer. In order to retain complexity large quantities of template as well as of vector and randomized primers are used, and product diversity is further ensured by pooling of multiple independent PCR-reactions. Subsequently, an overlapping PCR fragment containing the remaining vector segment is produced by standard PCR. Finally, this overlapping segment is joined with the PCR fragments containing the random DNA using another unique set of terminal primers.

DNA fragments produced by Taq DNA polymerase enzyme may contain additional nucleotides at the 3' DNA strand. These extra nucleotides will be deleterious for combining two PCR products with overlapping termini into one fragment. By addition of Klenow DNA polymerase enzyme these nucleotides can be removed by the 3' to 5' exonuclease activity of said enzyme, increasing the combining efficiency.

Alternatively the PCR product can be trimmed at the termini by digestion with restriction enzymes whose recognition sequences have been incorporated into the oligonu-

cleotides used as primers for the PCR reaction. By utilization of the temperature cycle method developed by (Lund et al., Nucleic Acids Res., 24, 800-801, 1996) the efficiency of the ligation reaction can be increased and the ligation product used for direct transfection of the packaging cells.

To maintain diversity, the PCR-generated linear vector DNA is used directly for transfection of packaging cells, and virions containing the pool of different vectors are harvested under transient conditions. Small bicistronic single transcript vectors containing a random peptide translation cassette followed by an internal ribosomal entry site (IRES) from EMC virus directing translation of a Neomycin (Neo) resistance gene. The Neo gene functions as a selection marker to allow titre determination and elimination of non-transduced cells, if necessary. Other available relevant vectors employ other selectable genes such as phleomycin and hygromycin B resistance and have the peptide translation cassette after the IRES element. Alternatively, even smaller vectors, carrying only the peptide expression cassette and lacking a selection marker can also be used.

The use of small vectors (below 4 kb of DNA) has the major advantage of allowing simple PCR-mutagenesis and amplification without ligation and cloning steps. Another important advantage of using small vectors is that the vectors in a pool of target cells can be amplified directly by PCR and retransfected into packaging cells, hence allowing multiple rounds of selection to remove time-consuming analysis of false positives or contaminating cells. Direct sequence analysis of the derived random plasmid clones is used to assure the randomness of the expression library. The standard vector capable of expressing the random peptide library is shown schematically in Fig. 1.

To maintain diversity of the vector pool a high fraction of the vector RNA transcripts must be encapsidated into retroviral particles. By transfection of the packaging cells with a DNA construct expressing a tRNA matching the corresponding PBS in the retroviral vector we can increase the production of functional vector containing virus particles 10 fold under transient conditions..

A new packaging cell line will be generated after a single transfection of a construct encoding all retroviral proteins. To diminish the risk of generation of replication competent virus and to obtain maximal expression the vector will have the following simplified outline: promoter-gag-pol transcript-IRES-phleomycin resistance gene-IRES-env-polyadenylation signal. One advantage of said vector is that the phleomycin resistance gene enables selection for high expression and that the sheer size of the vector transcript restricts the encapsidation into retroviral particles thereby diminishing the risk of generation of replication competent virus. The size limit for encapsidation of RNA transcripts is about 10 kb.

In addition to traditional packaging cell line a semi-packaging cell line with a corresponding minivirus-vector will be used. The semi-packaging cell line consists of vectors encoding two mutated gag-pol transcripts complementing each other. The use of two different gag-pol transcripts reduces the risk of generating wild type virus. Each cell in the semi-packaging cell line now contains all retroviral proteins needed for production of retroviral particles except the envelope proteins these proteins are supplied by the minivirus-vector. This vector is a bicistronic vector with following outline LTR-PBS-packaging signal-random peptide-IRES-env-polypurine tract-LTR. This vector will be able to transduce the semi-packaging cells as these do not produce envelope proteins prior to infection with the minivirus-vector.

Thus, infection of the semi-packaging cell will not be restricted by receptor interference.

5 Restrictions upon the "randomness" of the peptide sequence can be introduced, for the purpose of limiting the number of available sequences in the mammalian cellular library and for introduction of e.g. N-glycosylation sites, or other post-translational modifications of all expressed peptides if so desired. Purification tags -  
10 e.g. poly-His or others - can also be included in the expressed peptides for facilitating the purification and identification of the peptide itself. This is necessary for the identification of post-translational modifications, which were not obvious from the peptide sequence.

15 A population of eukaryotic cells is infected with the retrovirus carrying genetic constructs containing random DNA sequences which encode a library of millions of random peptides. Initially an excess number of virus compared to eukaryotic cells can be used. This leads to expression of a number of different peptides within each eukaryotic cell. These cells can subsequently be screened as described below, and the DNA can be isolated e.g. by  
20 PCR and used for reinfection of other cells. If an appropriate ratio between the number of retrovirus containing the random DNA sequences and cells is chosen, each cell will be transduced with a different random DNA sequence ("one cell - one ribonucleic acid or peptide"). This eventually enables the identification of an active peptide.  
30

The peptide may optionally be targeted to different compartments in the cell by incorporating appropriate signal sequences in the translated sequences.

**EXAMPLE 2**

*Identification of T cell epitopes by the use of mammalian intracellular expression libraries*

5

The interleukin 2 (Il-2) dependent cell line, CTLL-2, is transfected with the murine Major Histocompatibility (MHC) class I molecule,  $K^b$ . Subsequently this cell line is infected with a retrovirus peptide library which was described in Example 1. This CTLL-2 peptide library expresses a wide range of random peptides, and if appropriate  $K^b$  associated anchor residues known to be important for peptide binding to  $K^b$  are introduced in the retroviral peptide sequence, a large library of peptides bound to  $K^b$  are presented on the surface of the CTLL-2 cells.

10

$K^b$  restricted T cell hybridomas are generated against an appropriate virus antigen using conventional cellular immunological technology (Current Protocols in Immunology, Eds. Coligan et al., NIH). Such hybridomas secrete Il-2 upon recognition of antigen. A T cell hybridoma, which recognizes an unknown  $K^b$  bound virus T cell epitope, is subsequently incubated with samples of the  $K^b$  CTLL-2 library. If the hybridoma recognizes a peptide, the CTLL-2 cell presenting the peptide will be stimulated to proliferate by the Il-2 secreted by the hybridoma. In that way the CTLL-2 cell expressing the unknown virus T cell epitope can be selected and cloned. From this clone the DNA sequence encoding the peptide epitope in question can be isolated using PCR technology followed by conventional DNA sequencing. This eventually leads to identification of the unknown virus T cell epitope.

20

25

30

**EXAMPLE 3**

35

*Identification of biologically active peptides or ribonucleic acids which regulate cell surface expression of proteins*

Cells expressing the immunoregulatory membrane molecule, B7, are infected with the random peptide libraries constructed as described in Example 1. In this example a random eight-mer peptide library is introduced.

5

Using specific monoclonal antibodies the expression of B7 is analyzed by conventional methods. Cells which up- or down-regulate B7 can be selected either positively, e.g. using Fluorescence Activated Cell Sorting, or negatively, e.g. using appropriate antibodies in combination with lysis by complement.

10

Cells which show changes in expression of B7 are cloned by conventional means, and the DNA introduced by retroviral vectors is isolated using PCR and a set of retrovirus specific primers. The peptide sequence or possibly the RNA corresponding to said DNA may be able to modify the expression of B7 and hence the activation of T cells. This can subsequently be tested in conventional T cell assays.

20

#### EXAMPLE 4

*Identification of a F(ab) fragment capable of modifying the immunoregulatory molecule B7.*

25

A retroviral library encoding the variable heavy chain ( $V_H$ ) as well as the variable light ( $V_L$ ) gene fragments of the immunoglobulin molecule is produced. The gene region of both gene fragments corresponding to the antigen binding site of the resulting F(ab) fragments contains furthermore partly random gene sequences as described in example 1. This will lead to a large number of diverse peptide sequences in the antigen binding site of the F(ab) fragment. The retroviral vector library therefore encodes a large number of different F(ab) fragments with different antigen binding specification.

30

35



This library is subsequently transduced into a cell clone in such a way that each cell expresses a single F(ab) fragment species in e.g. the cytoplasm. Phenotypically changed cells are subsequently cloned and the sequence of the peptide in the intracellular (Fab) fragment is identified as described in example 1. This antibody can subsequently be produced in large scale by conventional means and be used for affinity purification of the cellular target protein responsible for the biological change of the cell phenotype. This can e.g. be done from lysates produced from the original non-modified cell clone.

Alternatively the retroviral F(ab) library can be constructed using e.g. a poly-His tag or other appropriate tags. In that way the antibody and the corresponding target can be isolated directly from the phenotypically changed cell by affinity chromatography. The isolated target can subsequently be identified by e.g. N-terminal amino acid sequencing in combination with conventional cloning methodology. Such target proteins are very important drug targets for further drug discovery.